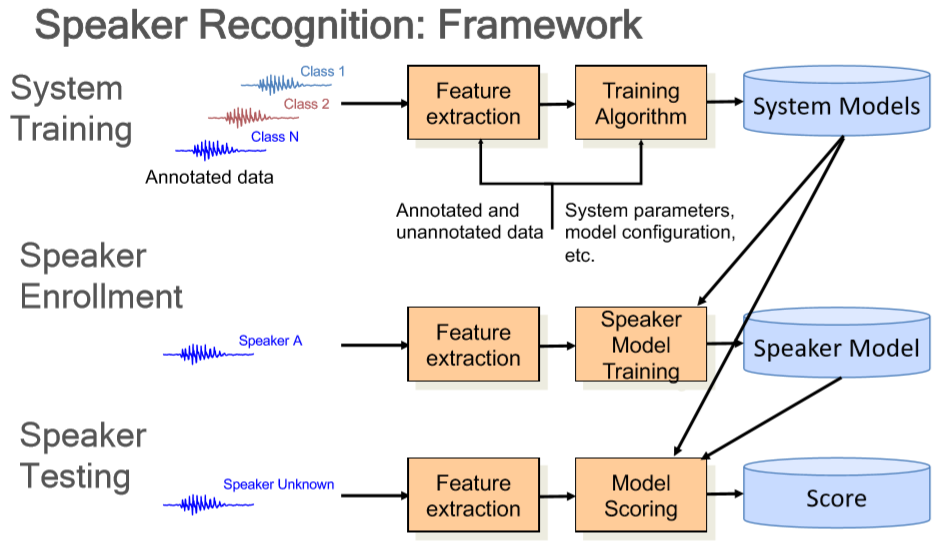
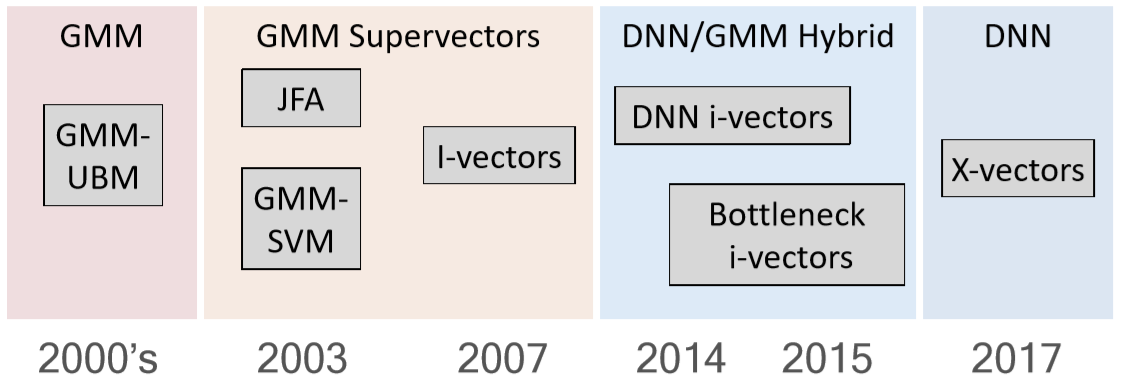
# 相关的问题

## ~~解决框架~~

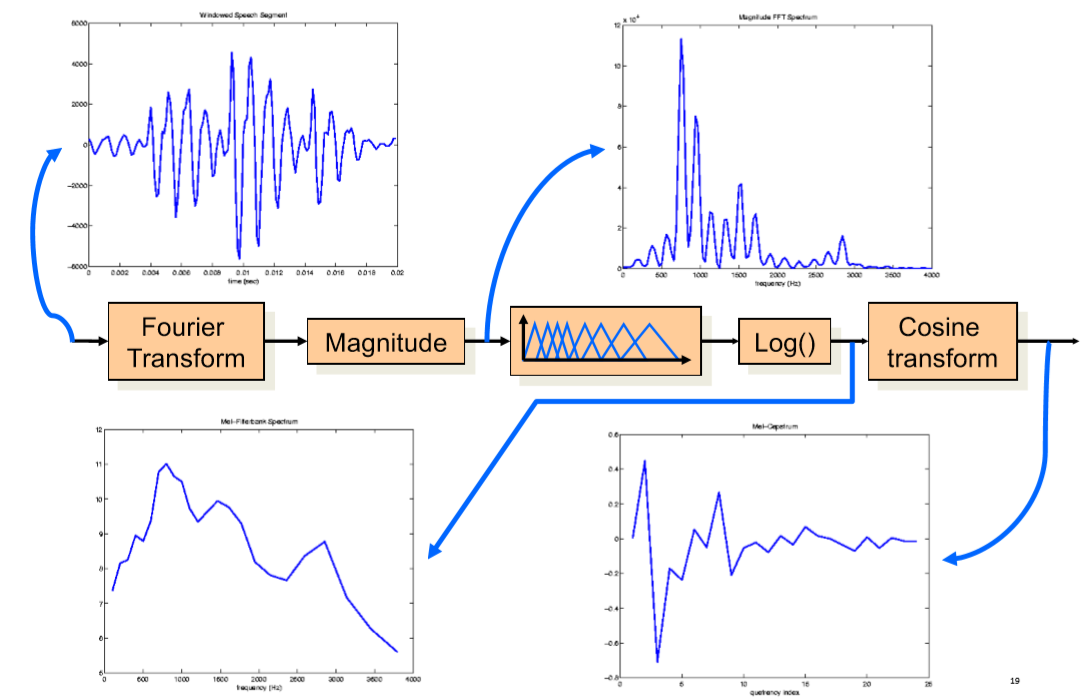


### 细节

## 方法的演进



## ~~音频数据的处理（梅尔倒谱）~~



* 傅里叶变换
* 放大
* 取对数log()
* 余弦变换

### 音频文件的存储细节

## 提取哪些特征

从低级到高级：

* 呼吸、鼻
* 韵律、节奏、语调、音量调节
* 语义、发音、特质

## 检测模型的时候如何排除内外因素的干扰？

* 外在
* 录音设备
* 声环境
* 传输的通道
* 内在
* 健康状况
* 紧张
* Role?

## 条件方面的挑战

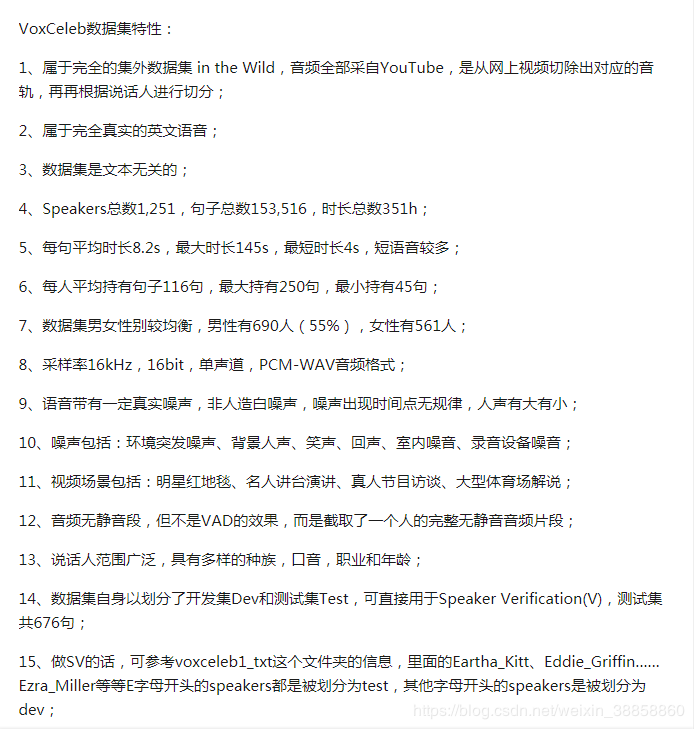
* 校准-确保系统在部署后符合预期的错误率
* 远距离语音-低SNR音频和混响
* 鸡尾酒会问题-可以以较高精度识别一个人所讲的话，但是当说话的人数为两人或者多人时，语音识别率就会极大的降低

## 数据（<http://www.robots.ox.ac.uk/~vgg/data/voxceleb/>）

VoxCeleb1&2：

全自动的数据采集

### 数据集的特点



## TensorFlow

* 版本选择

在 CPU 上运行 时，TensorFlow 本身封装了一个低层次的张量运算库，叫作Eigen；在GPU 上运行时，TensorFlow 封装了一个高度优化的深度学习运算库，叫作 NVIDIA CUDA 深度神经网络库（cuDNN）。

* 版本匹配（十分重要）

tensorflow 1.6.0对应keras 2.1.5

* Tensorboard出错

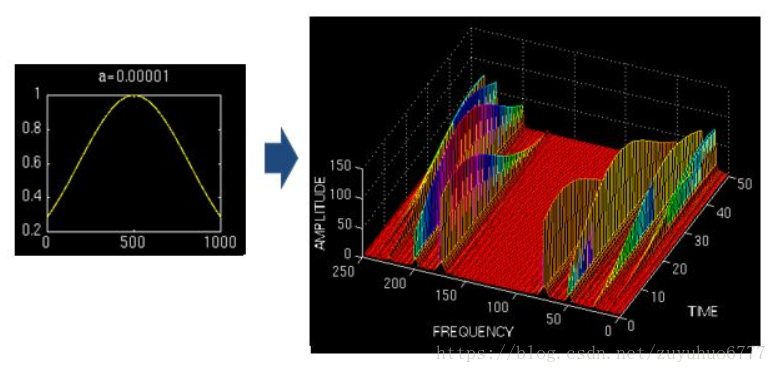
No attribute ‘SesionHook’

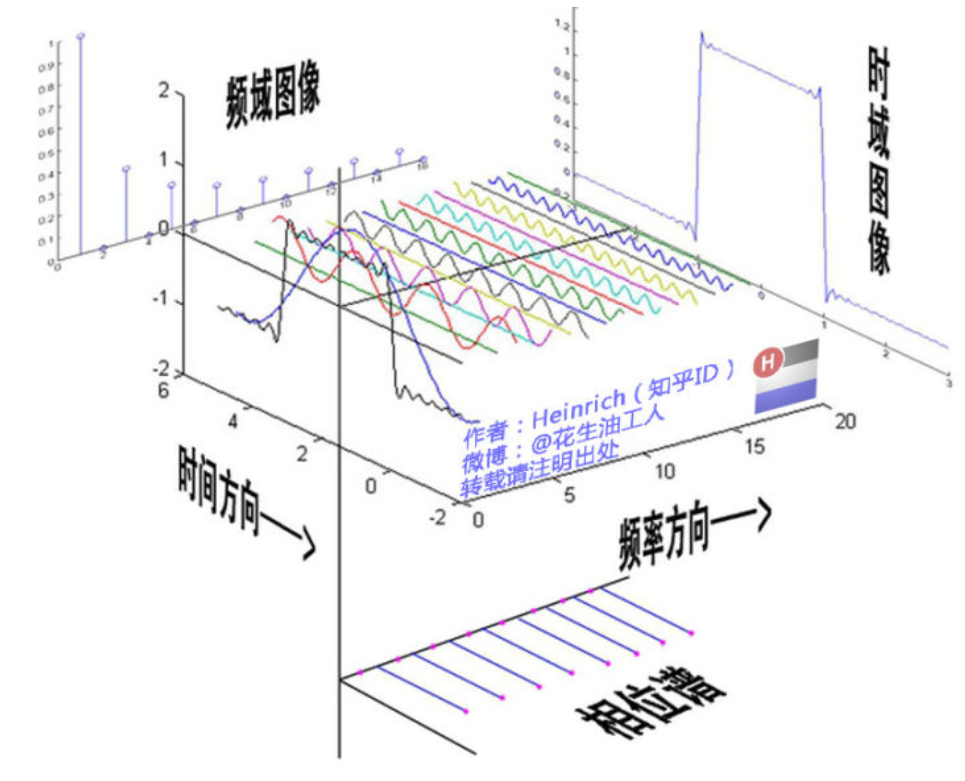
解决：

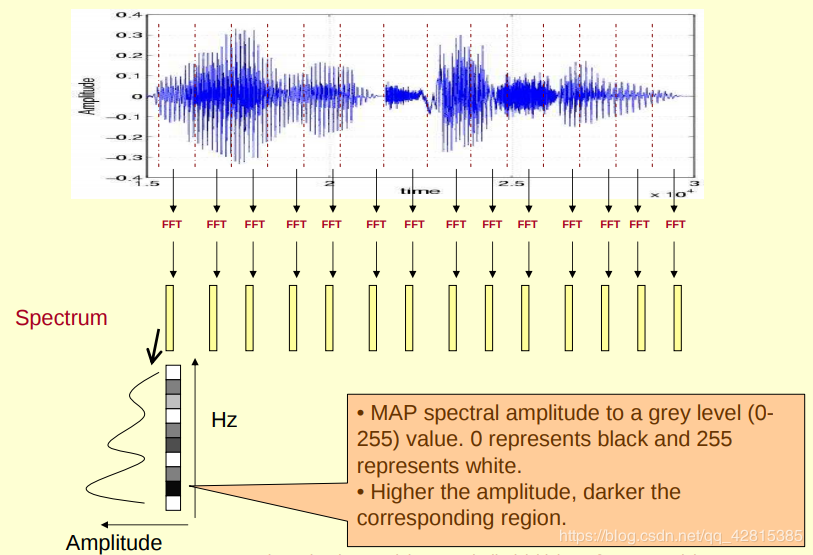
Pip uninstall tensorboard

pip install tensorboard==1.6.0rc0

## 短时傅里叶变化







Fn=(n−1)∗Fs/NFs(采样率)为16KHz，表示一秒钟内采样16000个点;N，表示对多少个点做FFT,如果一帧里面的点的个数小于N就会zero-padding到N的长度; \*每个点对应一个频率点，某一点n（n从1开始）表示的频率为Fn=(n−1)∗Fs/N，第一个点（n=1，Fn等于0）表示直流信号，最后一个点N的下一个点（实际上这个点是不存在的）表示采样频率Fs。

